

A cloud-based location assignment system for packaged food allocation in e-fulfillment warehouse

Yasmin YY Hui¹, KL Choy¹, GTS Ho¹, KH Leung², and HY Lam¹

Abstract

E-retailing is the delivery of products and services over the Internet, and it has been growing rapidly in recent years. Logistics services in the related warehouses have to be enhanced accordingly in order to handle the huge amount of e-orders in a shorter timeframe. Moreover, since the e-retailing of packaged food has been getting popular, these warehouses inevitably have to handle food products requiring special care, including temperature, humidity and insolation control, and expiry date management. Consequently, suitable storage guidelines have to be formulated for these warehouses to cope with their more demanding daily operations. A cloud-based location assignment system (CLAS) is thus proposed in this article, which provides timely and comprehensive solutions to the storage location assignment problem for warehouse operators. CLAS applies (i) cloud-based infrastructure for enhancing the agility of the proposed system, (ii) fuzzy association rule mining for predicting the potential storage time of various types of packaged food, and (iii) fuzzy logic and association rule mining for offering the most preferred storage locations of various products. A prototype of the CLAS is tested in a warehouse which involves packaged food handling. The test results show that the operational efficiency measured in terms of storage assignment decision-making time and order-picking time and the quality of the storage assignment solutions measured in terms of product disposal rate and product return/exchange rate have been improved in the case company.

Keywords

Storage location assignment problem, cloud-based decision support system, packaged food industry, e-fulfillment, fuzzy reasoning

Date received: 3 November 2016; accepted: 26 November 2016

Introduction

The trend of e-retailing has been adding pressure to packaged food warehouse operations. E-retailing has been thriving in recent years, as reflected by the annual growth rate of online shopping sales of 17% from 2007 to 2012, equivalent to US\$521 billion in 2012, and an estimated sales amount of US\$1248.7 billion in 2017.¹ The e-commerce sales in China increased by 40% from 2013 to 2014, where the growth was mainly contributed by the food sector.² Among various food products that are sold online, packaged food is a major category of food e-retailing due to consumer trust in fresh food products.^{2,3} Consumers have prior experience regarding packaged food, therefore, any perceived risk of buying

packaged food is lower.⁴ Therefore, packaged food has become a common option of online food purchase.

The warehousing of e-orders, however, is more complex and time sensitive than that of traditional orders.

¹ Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

² Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Pok Fu Lam, Hong Kong

Corresponding Author:

KL Choy, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
Email: kl.choy@polyu.edu.hk



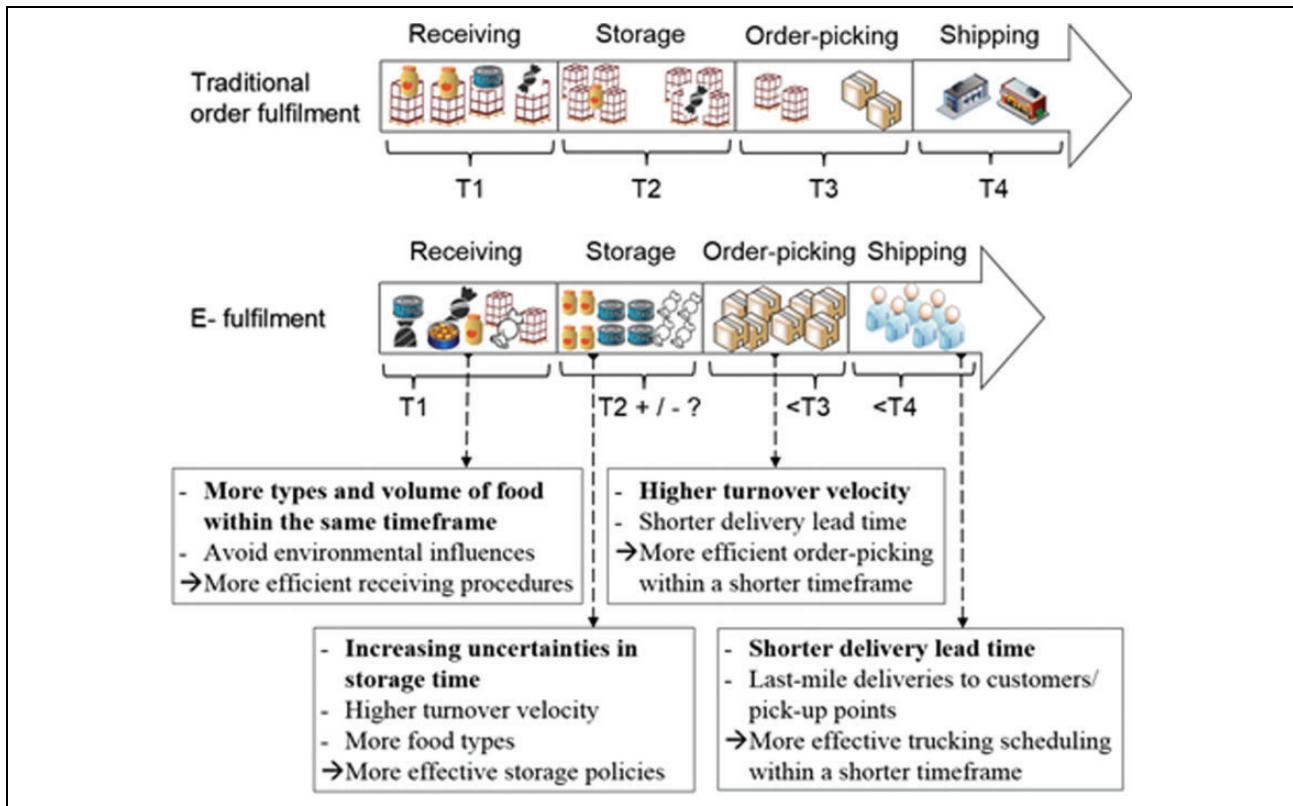


Figure 1. Operational challenges faced by decision makers of the e-fulfillment packaged food warehouse.

Comparing with fulfilling traditional orders, the operational challenges faced by the packaged food warehouse in fulfilling e-orders are shown in Figure 1.

Regarding the receiving operation, because the customers can access and place orders to a much larger number of online stores throughout the world via the e-retailing platform, there are more diversified types and volumes of food that need to be handled within the same timeframe T_1 ,² when compared with traditional order fulfillment. Therefore, to avoid environmental influences placed on the packaged food due to long waiting times at the unloading dock, more efficient receiving procedures are required.

Concerning the storage operation, increasing uncertainties in the storage time of the stored products are caused by the customer-oriented e-fulfillment practices, of which the end-customers are the decision makers of how and when the products are consolidated and delivered. Unlike the traditional order fulfillment practice in which the product outbound time can be scheduled in advance, the product outbound time facing the e-fulfillment warehouse is unplanned. In addition, a higher turnover velocity of products and more food types are observed in the e-fulfillment warehouse, due to the ease of placing orders by customers to any supplier at any time. This ordering convenience boosts smaller sized and larger numbers of orders that have to be delivered quickly. In order to cope with the more complex storage operations, more effective storage policies are thus needed.

In respect of the order-picking operation, it is affected not only by the phenomenon of the higher turnover velocity mentioned in the last paragraph but also the higher expectation of the customer, who perceives delivery lead time as a determinant factor of online shopping experience.^{5,6} Therefore, more efficient order picking is pursued within a shorter time frame.

Finally, the shipping operation becomes more time sensitive too. The expectation of shorter delivery lead times requires faster last mile deliveries to customers or pickup points; therefore, a more effective trucking scheduling within a shorter timeframe is needed.

In view of the operational challenges placed by e-retailing, the packaged food warehouse is in need of some new approaches to cope with the challenges. This article thus aims to propose a decision support system (DSS) to improve the operational performance of the packaged food warehouse. The proposed DSS should be able to (i) identify variables related to the storage time of the stock-keeping unit (SKU) through a variable selection (VS) technique, in turn to predict the range of SKU storage time and (ii) provide storage guidelines to the packaged food warehouse efficiently and effectively. In order to design an effective DSS, previous literature is studied in "Literature review" section. The design of the proposed system is then illustrated in "Cloud-based location assignment system" section and tested through a case study in section "Case

study.” In “Results and discussion” section, the system is validated through conducting a performance evaluation, and the final section gives the conclusions.

Literature review

In this section, the literature review covers four main areas. First, the specific operational problems that needed to be tackled in the packaged food warehouse are identified. Second, the applications of cloud-based DSS (C-DSS) are studied to assess its potential in enhancing the research methodology. Third, the data mining (DM) techniques that are commonly used in VS in the DSS are reviewed. Finally, the artificial intelligence (AI) and DM technologies that could be the core analytical techniques for storage location recommendation are reviewed.

Packaged food warehousing

The storage operations in a packaged food warehouse are particularly important for meeting the shelf life of food. Having suitable storage operations for packaged food means both the external storage factors such as humidity and the internal factors concerning the food such as nutrient content⁷ are maintained in the best conditions. Meanwhile, the study of packaged food e-retailing reveals that the packaged food warehouse has to equip itself for fulfilling a larger amount of the more time sensitive and unpredictable e-orders.⁸ Since the arrival time and volume of products become uncertain and the order-picking operations have to be more efficient, these require for a DSS that can assist the warehouse operator in making more timely and complex decisions regarding the storage operations.

Among the decisions involving storage operations, the storage location assignment problem (SLAP) requires a complex decision to be made that directly affects the quality of food products and the efficiency of the receiving, storage, and order-picking operations.^{9,10} SLAP is about the assignment of incoming products into storage locations.¹¹ The existing studies in SLAP cannot meet the needs of a packaged food warehouse. They mainly aim to enhance operational efficiency and optimize the space utilization in the warehouse,¹² but seldom take the well-being of products into account. However, while other products are less susceptible to the storage environment, the shelf life of packaged food is greatly affected by the environment. To be specific, Pereira de Abreu et al.¹³ stated three factors that would affect the shelf life of food. They are materials and types of food package, internal characteristics of food such as water and nutrient content of the food, and external storage environment such as storage temperature and relative humidity. Without a proper storage location, the packaged food would deteriorate or be damaged easily. In view of the importance of being assigned to a proper storage location to the packaged food, designing an SLAP DSS that considers both the food quality and

operational efficiency of a packaged food warehouse can therefore be a contributive research area for the industry.

Cloud-based DSS

DSS is a class of system that assists decision makers during the problem solving process by retrieving data and testing alternatives.¹⁴ It is becoming popular for industrial practitioners to host part or all of their DSS in a cloud environment, in order to enjoy the low-cost scalability and flexibility offered by the cloud.¹⁵ A C-DSS is a component-based system that is reusable, substitutable, scalable, and customizable.¹⁶ Since the system proposed in this research aims to serve warehouses that need to fulfill e-orders, cloud computing can facilitate data collection from a large number of parties involved in e-transactions. Furthermore, as the demand derived from e-retailing can be fluctuating and unpredictable, the scalability of the cloud infrastructure offers flexibility to the warehouse operator to scale the computing services up or down, according to actual needs. The characteristics of C-DSS make it promising to cope with the dynamic e-retailing environment.

In addition, because of the increasing number of e-retailing operations, such as e-shopping and e-payment,¹⁷ various cloud-based systems emerge to support the operations.^{18–21} Recent examples of cloud-based system application demonstrate the value and reliability of adopting cloud infrastructure in these supportive systems. Guo et al.¹⁸ illustrated how to make use of the information sharing function of a cloud-based decision-making system for better decision-making in supply chain coordination. Mann and Kaur¹⁹ designed a cloud-based DM framework in a clinical DSS that supports the security in cloud application. The system makes use of access control and information encryption functions of the cloud infrastructure to secure the sensitive personal information of patients. Dong et al.²⁰ proposed a multicloud-based evacuation system which shows that the stability of cloud infrastructure can be secured using the snapshot function of the cloud infrastructure. Demirkan and Delen¹⁶ also stated that data availability in cloud environment can be maintained by data replication. In summary, the aforementioned characteristics of C-DSS and the successful applications of C-DSS in other areas show that cloud computing is an essential and reliable component of an SLAP DSS for packaged food warehouse.

VS for storage time prediction

VS is a technique for selecting the relevant variables to develop a learning model.²² This technique is particularly useful in developing a DSS, because with more relevant input variables, the prediction performance, thus the recommendations offered by the DSS, can be enhanced.²³ For example, Ząbkowski and Szczesny²⁴ reduced the number of variables from 205 to 26 after going through the VS procedure, rendering a more accurate and simpler

prediction model in customer insolvency. Elwakil and Zayed²⁵ apply VS methods to find relation between quantitative and qualitative variables from the project data in order to predict the work task durations. These applications of VS demonstrate the value and function of VS that are usable in this study. When designing a DSS for SLAP handling, the storage time of the SKU is a piece of crucial information, resulting in a number of studies on the relationship between demand variability and SKU storage time.^{26,27} However, there are factors other than demand variability that potentially affect the storage time; therefore, the proposed DSS should execute VS to identify the most relevant factors and in turn predict the range of storage time more accurately and efficiently.

Regarding the approaches of VS, they can be implemented by statistical and DM techniques. According to the study by Nettleton,²⁸ statistical approaches are generally used to find the relationships among the variables and in looking for linear relationships between them. In order to use statistical models such as logistic regression, strict assumptions on normality and linearity, and requirements in sample size have to be complied with. These restrictive prerequisites made statistical models difficult to be applied.²⁹

On the other hand, DM approaches are applicable in unveiling nonlinear relationships among variables and are effective in analyzing the relationship between input variables and the specific output variables that are determined by the analyst beforehand.^{28,30} In this research, since the input variables do not necessarily have linear relationships among themselves and the output variable is specified as the storage time, DM approaches are thus preferable.

Comparing with previous studies that applied clustering and rule induction as the DM techniques for VS,^{30–32} fuzzy association rule mining (FARM) is more preferable in the context of this study.³³ FARM is able to cultivate relationships between ranges of the parameters, instead of only the relationships between Boolean attributes.³⁴ This additional functionality of FARM enables the description and the specific happening conditions of the if–then relationships between the input and output parameters.^{34,35} For example, while the other techniques only suggest that “sales turnover rate is greatly positively related to the SKU storage time,” FARM can further describe the situation such as “IF sales turnover rate is high, THEN the SKU storage time is short.” The range of storage duration, which is a crucial input variable SLAP handling, can thus be accurately predicted by FARM.³⁶ Therefore, FARM is a suitable technique for identifying the factors and patterns concerning the storage time of SKUs, in turn assisting decision support in SKU allocation in the packaged food industry.

SLAP decision support

Apart from VS, the main function of the proposed system is to recommend the most suitable locations for various

products in a packaged food warehouse. Since part of the important attributes concerning packaged food are difficult to be measured by crisp values, such as the vulnerability of food package and the resistance ability to the external storage environment, fuzzy logic (FL) should be adopted as the core analytical tool of the proposed system. FL is able to convert vague information into numerical values that fall between 0 and 1 of the membership levels. This functionality of FL allows the conversion of human reasoning into precise rules that can be processed by the DSS.^{37–39} FL has been applied in SLAP handling in previous studies,^{38,40,41} which proved that it can be a promising tool for the proposed system.

Apart from FL, association rule mining (ARM) is another effective technique that provides quality decision support in SLAP. ARM is a DM technique that aims to unveil the co-occurrence relationships among data sets in an efficient way.⁴² It has been applied in SLAP through finding the ordering patterns of SKU, in turn locating SKUs that are frequently delivered together into closer locations.^{11,43,44} Since customer orders in the packaged food industry often involve multiple products, ARM can be a suitable tool to reveal the hidden ordering patterns. Therefore, it can be embedded into the proposed DSS, where deemed appropriate, in order to further enhance the quality of the storage location recommendation.

The implications of the literature review are four folds. First of all, the concerns of food safety and e-retailing facing the packaged food industry emphasize the importance of having an SLAP DSS in the packaged food warehouse. Second, the scalable cloud infrastructure could be useful for the packaged food warehouse that needs to deal with dynamic number of e-orders and partners. Third, FARM is a promising technique for VS of the storage time of SKU as it provides more indicative information. Finally, FL and ARM are the suitable analytical techniques to be applied in the proposed DSS, because FL is particularly good at analyzing the ambiguous storage requirements of food, while ARM is effective in revealing ordering patterns hidden in customer orders. The four implications are then used to design the proposed system.

Cloud-based location assignment system

The system architecture of a cloud-based location assignment system (CLAS) is shown in Figure 2. The system is comprised of three modules:

information consolidation module, FARM variable selection module (VSM), and decision support module (DSM).

Information consolidation module

The data sources of CLAS are categorized into the internal data stored in the corporate's system, such as enterprise

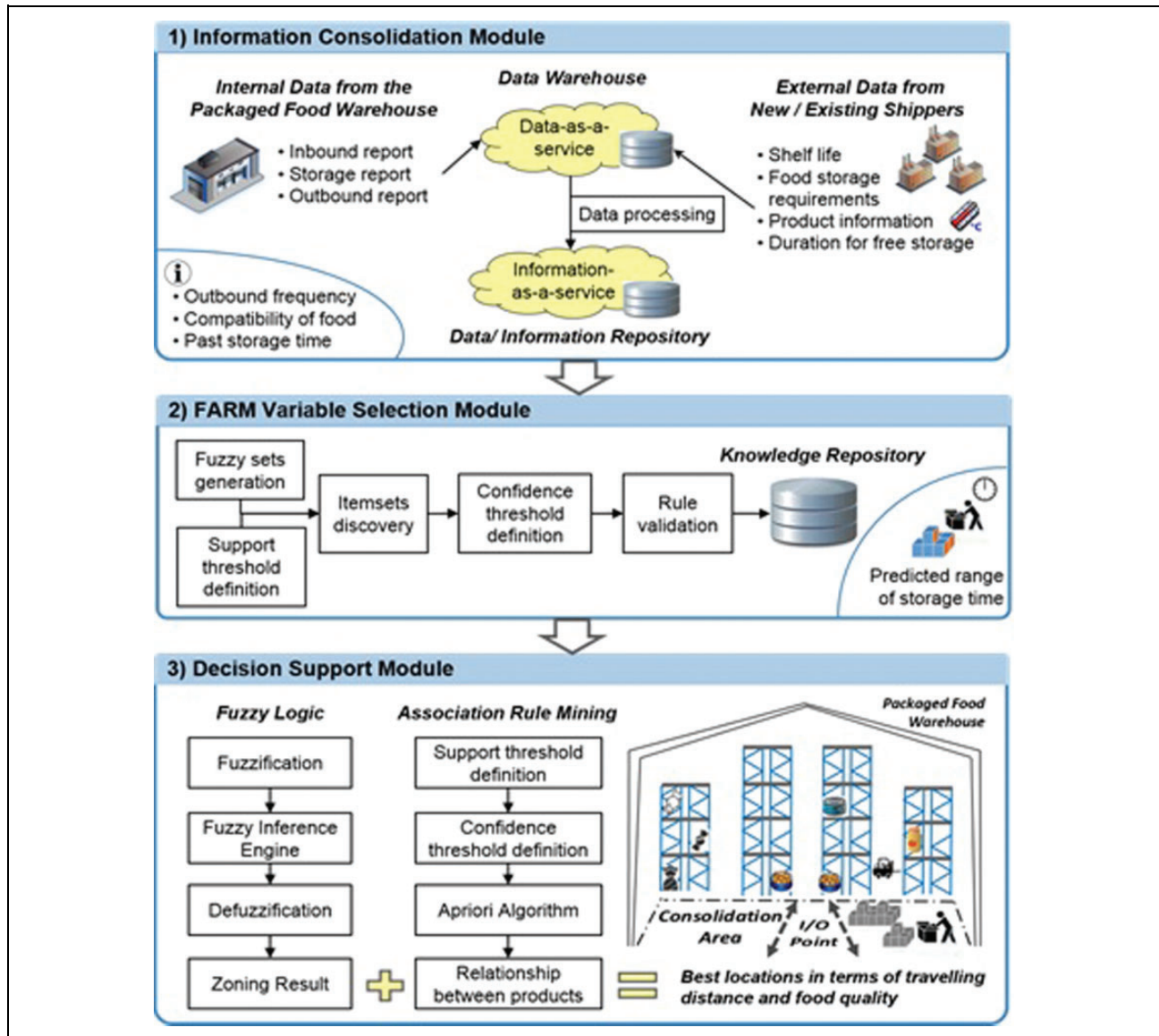


Figure 2. System architecture of CLAS. CLAS: cloud-based location assignment system.

resources planning (ERP) and warehouse management system, and the external data provided by business partners, such as shippers from various regions that deliver their products to the warehouse for preparing the e-orders. The data obtained are then stored in a cloud-based repository that can store both data and information. Data preprocessing and processing are implemented in the cloud infrastructure in order to produce usable data and information for further analysis. The potentially useful information includes outbound frequency, past storage time of SKUs, and the food compatibility.

FARM VSM

The usable data and information available in the cloud repository are then analyzed with different DM and AI techniques in order to provide different decision support

functionalities. VSM is responsible for selecting relevant variables for the attribute “range of SKU storage time” and at the same time predicts the range of SKU storage time. To begin with, parameters that potentially affect the SKU storage time are identified through studying the historical data and interviewing domain experts, who are decision makers of the packaged food warehouses. The fuzzy sets of each parameter are then generated through defining the membership functions and universe of discourse of each parameter. After calculating the fuzzy counts of the parameters, the support count threshold can be set. Afterward, itemsets are discovered through executing the mining algorithm. Meanwhile, the confidence threshold is defined to filter out itemsets that do not have a sufficient number of occurrences in order to obtain the potential rules. Finally, the potential rules have to go through a rule validation process. Since Jain and Srivastava⁴⁵ state that the performance of

the DM-based predictive model is evaluated by its ability to predict accurately with new data, the rules are thus validated by checking the number of cases in which their predicted range of storage time mismatches the actual range of storage time. The rules that obtain an acceptable accuracy are stored into the knowledge repository to provide the predicted range of storage time to the next module.

Decision support module

As the predicted range of SKU storage time is obtained from the last module, the next step is to form the knowledge base for completing the location assignment decision support function of CLAS. DSM consists of two main submodules, fuzzy zoning and ARM, which apply FL and ARM, respectively. The function of the fuzzy zoning submodule is to suggest the warehouse zone to be assigned to the SKUs and the attention needed when handling different SKUs. This submodule has three main steps: fuzzification, fuzzy inference engine analysis, and defuzzification. Fuzzification is carried out after determining the membership function and universe of discourse of the input variables, while the fuzzy inference engine analysis is done using knowledge of the packaged food handling. Such knowledge is expressed in if-then rules and is retained in the inference engine.

While the fuzzy zoning submodule may suggest the same warehouse zone for a wide range of SKUs, the SKUs that are assigned to the same zone could have interrelationships that further determine the closeness of their locations. Therefore, the ARM submodule emerges to find out the interrelationships among the SKUs through cultivating association rules from the ordering records. The rule mining process is initiated by setting the minimum support and confidence threshold and is completed through the a priori algorithm. The revealed relationships between SKUs orders can enhance the storage location assignment decision, in turn further improving the order-picking efficiency.

Case study

A case study is initiated in ABC Company (ABC, Hong Kong) to test the feasibility of CLAS. ABC is a Hong Kong-based trader of packaged food located in San Tin, Hong Kong. It has a warehouse in Hong Kong situated close to the border with China that deals with transshipment of packaged food. In order to expand its business, ABC started to set up an e-retailing platform. However, its current workflows are not capable enough to handle the more diversified types of food and the more time-sensitive order-picking operations brought by e-fulfillment. The problems facing the warehouse can be summarized as (i) a lack of a proper storage policy for quickly and suitably assigning the storage locations for the SKUs and (ii) a lack of a mechanism for predicting the storage time of products, given that the outbound time of SKUs is becoming unpredictable

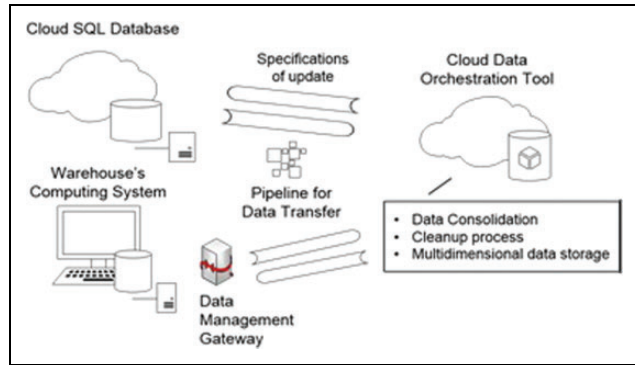


Figure 3. Data flows concerning the C-DOT. C-DOT: cloud data orchestration tool.

under e-retailing. The first problems could lead to food deterioration if the product receiving time is too long. In view of the problems and the possible consequences, CLAS is proposed and implemented in ABC to improve the food storage operation and the operational efficiency in the warehouse.

Three stages are involved in the implementation of CLAS: (i) construction of cloud-based database (C-DB), (ii) implementation of VSM, and (iii) implementation of DSM.

Construction of C-DB

Regarding the construction of the C-DB, the cloud SQL database service is chosen by the case company as it is fully compatible with the on-premises MS office and SQL server. The data sources related to CLAS are the Excel files, ERP, and SQL database found in the warehouse's computing system. Besides, shippers provide the Excel files containing the shipment information to ABC. All data stored in an on-premise system has to be transferred to the C-DB through the interface of the cloud service provider.

After storing the data into the C-DB, a cloud data orchestration tool (C-DOT) is employed. As shown in Figure 3, the C-DB is linked to the C-DOT directly while the local computer system and C-DOT are connected through a data management gateway. The data transfer is executed through the data pipelines set between C-DOT and C-DB, and C-DOT and local computing system of the warehouse. The settings of the pipeline determine the specifications of the update activities between the databases and the C-DOT. The C-DOT is responsible for data consolidation, data clean-up, and multidimensional data storage. Information such as sales rate, storage time, disposal rate, storage condition and package material of different kinds of packaged food should be ready for applying FARM, FL, and ARM in the later modules.

Implementation of VSM

While the database is ready, VSM can be implemented to obtain the predicted range of SKU storage time. There are

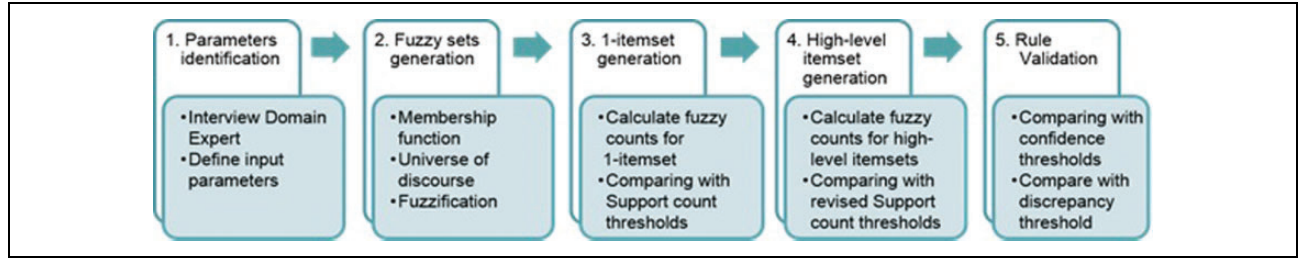


Figure 4. The five steps of VSM implementation. VSM: variable selection module.

five main steps involved in generating reliable fuzzy association rules that describe the if–then relationships between the input parameter and the output parameter, that is, the range of SKU storage time. The five steps are summarized in Figure 4³⁶ and are explained as follows:

Step 1: Parameters identification

In order to initiate the rule mining process, parameters have to be defined. A domain expert in ABC is consulted to obtain the parameters that possibly relate to the SKU storage time. Six parameters are identified after interviewing the domain expert. The six parameters, the symbol representing the parameters, the calculation methods, and the range of the parameters are summarized in Table 1. Parameters A to E are input parameters that affect the output parameter F to different extents.

Step 2: Fuzzy sets definition

After defining the parameters, the fuzzy sets of each parameter have to be set. The historical data and expertise of the domain expert provide necessary information for setting the membership function and universe of discourse of each parameter. The fuzzy set of the parameters can then be set based on the membership function and universe of discourse. The linguistics terms are also the fuzzy classes of a parameter that describe the fuzzy characteristics of the parameter. When all fuzzy sets concerning the parameters are defined, the itemset can be generated.

Step 3: One-itemset generation

In order to generate itemsets, two main procedures are involved. They are the calculation of fuzzy counts for one-itemset and comparing with the support count threshold. To calculate the fuzzy count, numerical values of previous records are first converted into membership values. Afterward, summations of the membership values of every fuzzy class, concerning each parameter, are done. Fuzzy counts can then be calculated after adding up the membership values of each fuzzy class. When the fuzzy counts are available, they are compared with the support count threshold to filter out itemsets without a sufficient number of occurrences. Six thresholds for the six parameters are obtained through experimenting with the fuzzy counts

Table 1. Information regarding the six parameters.

Parameter	Calculation method	Symbol	Range
Order size	Monthly average order quantity of the SKU for each order	A	0–6 pallets
Sales velocity	Average sales quantity to average receiving quantity of the SKU in a month	B	0–100 %
Expiry date	The number of months between the SKU receiving date and the expiry date of the SKU	C	1–24 months
Closeness to holidays	The number of days between the SKU receiving date and closest holiday	D	1–60 days
Past storage time	The average number of days between the receiving date and the outbound date of the same SKU in past 30 days	E	1–50 days
Actual storage time	The number of days between the receiving date and the outbound date of the current lot of SKU	F	1–50 days

SKU: stock-keeping unit.

concerning each parameter. When fuzzy classes contain fuzzy counts that are equal or greater than the designated support count thresholds, the classes can be kept to represent the parameters and be referred to as the one-itemset. However, when there are more than one fuzzy classes under the same parameter that pass the threshold, the class with the greatest fuzzy count value would represent the parameter. After the whole process, there should be six fuzzy classes (one-itemset) representing the six parameters for further processing in the next step, that is, high-level itemset generation.

Step 4: High-level itemset generation

The high-level itemsets are various combinations of the one-itemset, where two-itemset contains two one-itemsets, three-itemset contains three one-itemset, and so on. A high-level itemset is the original form of the fuzzy association if–then rule. To generate a high-level itemset, there are two main procedures which are the same as that for generating

Table 2. Formulas concerning the rule validation process.

Formula number	Function	Formula
1	Calculate confidence value of two-itemset	$\frac{\text{Fuzzy count of the two-itemset}}{\text{Fuzzy count of the first one-itemset involving the two-itemset}}$
2	Calculate confidence value of three-itemset	$\frac{\text{Fuzzy count of the three-itemset}}{\text{Fuzzy count of the first two-itemset involving the three-itemset}}$
3	Calculate discrepancy rate of rule	$\frac{\text{Number of mismatched records concerning the target rule}}{\text{Total number of records concerning the target rule}}$

one-itemset, except that the calculation methods are different. To obtain fuzzy counts concerning each high-level itemset, the smallest membership values of all records involving the itemset are added. The support count threshold of a high-level itemset is acquired using the largest support count threshold among the thresholds involving the itemset. When both the high-level fuzzy counts and support count threshold are available, they are compared with each other for filtering of the itemsets. The high-level itemsets that contain fuzzy counts equal or greater than their thresholds would be maintained as potential fuzzy association rules. Step 4 is repeated to generate itemsets that contain more and more numbers of fuzzy classes, until the fuzzy counts of the higher level itemsets are smaller than their thresholds.

Step 5: Rule validation

Step 4 is executed on the active items in ABC, where the active items are the SKUs that have been ordered and delivered in the past 2 months. Five hundred and fifty-three potential rules, involving more than 70 SKUs, have been cultivated and expressed in both two-itemsets and three-itemsets. However, since only the rules that contain parameter *F actual storage time* are valuable to the DSS, only 268 rules that involve parameter *F* are kept. The 268 rules are then validated by two thresholds, namely, the confidence threshold and the discrepancy threshold, in order to generate fuzzy association rules that can more accurately predict the range of SKU storage time.

Confidence threshold is the probability of the occurrence of the high-level itemset when its involved one-itemset occurred. Discrepancy threshold is the allowance percentage of mismatched records concerning a rule that reflects the prediction accuracy of the rule. For example, as predicted by the rule “A: medium and B: high,” when parameter A of an incoming batch is medium, the parameter B of the same batch would be high. A 10% discrepancy threshold indicates that given 10 records involving medium for their parameter A, only one of them is allowed to involve medium or low, instead of high, for its parameter B.

To validate the rules with the thresholds, the confidence threshold is compared against the confidence value of a rule, and the discrepancy threshold is compared against the discrepancy rate of a rule. The formulas for calculating the confidence value and discrepancy rate of rule are presented in Table 2. The confidence threshold and discrepancy

threshold are set to be 90 and 20%, respectively, after considering the results of experiments done with different threshold values and the opinions of ABC’s decision makers. After filtering out the rules that contain confidence values and discrepancy rates smaller than the two thresholds, 221 rules remain and are stored into a knowledge repository. Among the 221 rules, sales velocity and order size are involved in 72 and 6% of the rules, respectively. These results indicate that the most and least relevant parameter to storage time is sales velocity and order size, respectively. Five examples of the validated rules are shown in Table 3. The range of storage time of different SKUs expressed in short, medium, and long is provided by the rules in this module. Short, medium, and long storage time refer to 0–25, 12.5–37.5, and 37.5–50 days, respectively. This important piece of information is then used in the fuzzy inference engine in DSM for determining storage locations.

User interface of VSM. The specifications concerning the FARM algorithm can be edited by the user through the user interface of VSM, as shown in Figure 5. Through the interface, the user can specify the range of data in terms of date, the parameters involved, and the support count, confidence, and discrepancy thresholds for rule filtering. There are two rule generation boxes where the upper one generates rules without considering the discrepancy allowance threshold, while the lower one generates rules considering the discrepancy allowance threshold. The decision of using which rule generation box is made by the decision makers, who decide whether they want more rules or more accurate rules.

Implementation of DSM

Since the range of SKU storage time is acquired in VSM, DSM can be developed. Three main steps are involved in the development of DSM. The steps and their major procedures are shown in Figure 6 and are explained as follows:

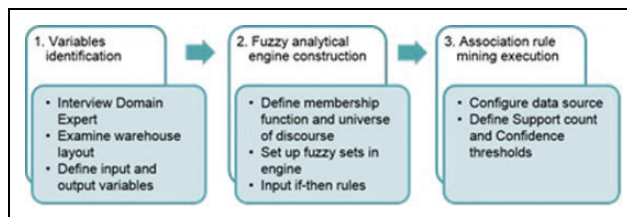
Step 1: Variables identification

In order to carry out the fuzzy zoning decision support function of CLAS, variables that will be processed by the fuzzy analytical engine have to be defined. While the input variables can be recognized by interviewing the domain expert in the company, the output variables are set according to the warehouse layout of ABC. The input and output variables defined are listed and described in Table 4.

Table 3. Examples of the validated fuzzy association rules.

SKU	Validated fuzzy association rules	Confidence value	Discrepancy rate (%)
T30	IF—Sales velocity is high, closeness to holidays is near, THEN—Actual storage time is short.	1.0	12
Whole almond 3lb	IF—Sales velocity is high, past storage time is short, THEN—Actual storage time is short.	0.98	11
Peanut butter 1360 g × 2	IF—Expiry date is medium, closeness to holidays is near, THEN—Actual storage time is short.	0.93	14
Chocolate assortment	IF—Sales velocity is medium, closeness to holidays is near, THEN—Actual storage time is medium.	0.91	19
Oats 4530 g	IF—Sales velocity is low, expiry date is long, THEN—Actual storage time is long.	0.91	16

SKU: stock-keeping unit.

Figure 5. User interface of VSM. VSM: variable selection module.**Figure 6.** The three steps of DSM implementation. DSM: decision support module.

Step 2: Fuzzy analytical engine construction

After the variables are identified, the fuzzy analytical engine is constructed with the fuzzy toolbox of MATLAB® and Simulink® Release 2008a (© 1984–2008 The MathWorks, Inc, MA, USA). This toolbox is used because of its

capability of integrating with the existing operation system in the company. The engine is comprised of fuzzy sets of each variable and the if–then rules that specify the relationships between the variables. Fuzzy sets are formed by the membership function and universe of discourse, which can be defined through studying the historical data regarding the variables and learning from the domain expert. After defining all fuzzy sets, the membership function, universe of discourse, and fuzzy sets are set up in the engine.

Afterward, the if–then relationships linking up the input and output variables are inputted into the engine. Based on the combinations of various membership functions of the first five input variables, 324 rules can be derived. In addition, another group of 324 rules is also derived, and the only difference between two groups of rules is that the fifth input variable is *sales velocity* instead of *storage time*. With this

Table 4. Input and output variables of the fuzzy analytical engine.

Type of variable	Parameter	Description	Range
Input variable	Tolerance to temperature change	The extent of resistance to temperature change in the storage environment	1–15°C
	Package vulnerability	The score given to an SKU according to the material and shape of the SKU's package	1–18 point
	Disposal rate	The percentage of items that are damaged in an incoming lot	0–0.01%
	Inbound size	The number of pallets involving the same SKU that arrives in the same batch	0–8 pallets
	Storage time	The predicted storage time of the inbound product generated from VSM	1–50 days
	Sales velocity	Average sales quantity to average receiving quantity of the SKU in a month, a complementary variable to storage time. It is only considered in the analysis when storage time is not available.	0–100%
Output variable	Warehouse zone	Virtual zones of the warehouse that have different extent of environmental variance and traveling distance to the warehouse entrance	1–5
	Vertical level	The lower or the upper deck of pallet	1–2

SKU: stock-keeping unit; VSM: variable selection module.

setting, the total of 648 rules can ensure that every SKU can be analyzed with the fuzzy inference engine. All the rules are inputted and stored in the fuzzy inference engine to form a knowledge repository. When the rules are available, the engine can perform the fuzzy zoning decision support function. The user of CLAS, however, would perform the analysis through the user interface of DSM.

User interface of DSM. The user interface of DSM is shown in Figure 7. Through the interface, the user can retrieve information regarding the incoming batch of products. After verifying the attributes of the product, allocation guideline can be generated in the box located at the upper right corner of the interface. The settings in the fuzzy analytical engine can be modified through retrieval from the interface. In addition, if the warehouse operators want to obtain more comprehensive allocation guidelines, which include the relationships between SKUs, they can use the box located at the lower right corner of the interface. This box initiates the ARM process, which is explained in the next step.

Step 3: ARM execution

Although the fuzzy analytical engine recommends a zone for SKU allocation, the traveling distance involved in order picking can be further reduced by executing ARM. The MS SQL Server DM Add-in (© Microsoft Corporation, Washington, USA) is employed to implement the rule mining, due to its easiness in use and compatibility with the MS SQL Server and MS office software (© 2012 Microsoft Corporation, Washington, USA) that ABC is using. Since ABC has installed the Microsoft SQL Server 2012 in their local computer, a free application of the Microsoft® SQL Server® 2012 DM Add-in is feasible. The two major procedures in executing ARM through the add-in are to configure the data source and to define the support count and confidence thresholds.

The data source of ARM can either be the sales report generated from the ERP or the transferred sales report extracted from the C-DB. After installing the add-in, the database in the SQL server is connected to MS Excel, and then the sales data in the database can be imported into the Excel sheet. After data importing, the “associate wizard” is initiated, and the minimum support and confidence thresholds are set in the tool to determine the strictness of rules generation.

Experiments are done to define suitable thresholds. Since the major objective of executing ARM is to identify as many combinations of the itemsets as possible, in turn to offer recommendations on storage locations for as many items as possible. The threshold definitions thus have to result in a relatively large amount of rules with acceptable confidence level. Five trials with different combinations of thresholds are done and shown in Table 5. The results of the trials indicate that the confidence threshold brings great impact on the number of rules generated, while the support count threshold brings great influence on the number of itemsets. Since even the unpopular items, that is, the items with fewer turnover, should be well located too, the support count threshold can thus be relatively loose. At the end, trial five offers preferable results: having a larger number of rules (26 rules) that comply with an acceptable confidence threshold (75%) and a large number of itemsets. The suggested thresholds for support count and confidence thresholds are thus five items and 70%, respectively.

After setting the thresholds, rules can be generated. As shown in Figure 8, there are 60 rules that comply with the two thresholds. The rules indicate which items are sold with which other items in the same transaction. For example, a type of fish oil and Docosahexaenoic Acid (DHA) product is always (confidence = 100%) sold with a type of calcium supplement (the highlighted rule). The rules are then exported from the rules browser to an Excel sheet for further processing and records.

The implementation of CLAS in ABC proves that CLAS is a feasible SLAP DSS that can provide allocation

Figure 7. User interface of DSM. DSM: decision support module.

Table 5. Results of experiments for thresholds definition.

	Support count threshold (in number of counts)	Confidence threshold (in %)	Number of rules generated	Number of two-itemsets generated	Number of three-itemsets generated
Trial 1	10	70	10	49	3
Trial 2	8	80	14	75	7
Trial 3	6	90	5	148	42
Trial 4	4	90	7	302	225
Trial 5	5	70	26	302	225

guidelines to packaged food warehouse operators. The performance of the system, however, is further evaluated in the next section.

Results and discussion

After presenting the details of CLAS implementation in the case company, the contributions and costs of the system have to be reviewed. The system is evaluated in regard to two aspects: benefits of CLAS application and cost analysis of CLAS implementation.

Benefits of CLAS application

Four key performance indicators (KPIs) were set in response to the purposes of developing CLAS. The KPIs were recorded before and after implementing CLAS and are shown in Table 6.

Considering the KPIs records and the number of valid fuzzy association rules generated by VSM, as explained in “User interface of VSM” section, CLAS implementation is proven to be beneficial to a packaged food warehouse in three aspects:

- i. *Provision of storage guideline for efficiency improvement*

The improvements that CLAS brought about to the efficiency of the warehouse are supported by the KPIs records of “SKU allocation time” and “order-picking time.” The average time spent in allocating SKUs dispatched from a container into the proper warehouse sections sped up by 11.8%. This improvement indicates that CLAS can serve its purpose of providing decision support to stock keepers who make the SLAP decision. Meanwhile, the average time spent in picking an order was reduced by 25.5%, suggesting

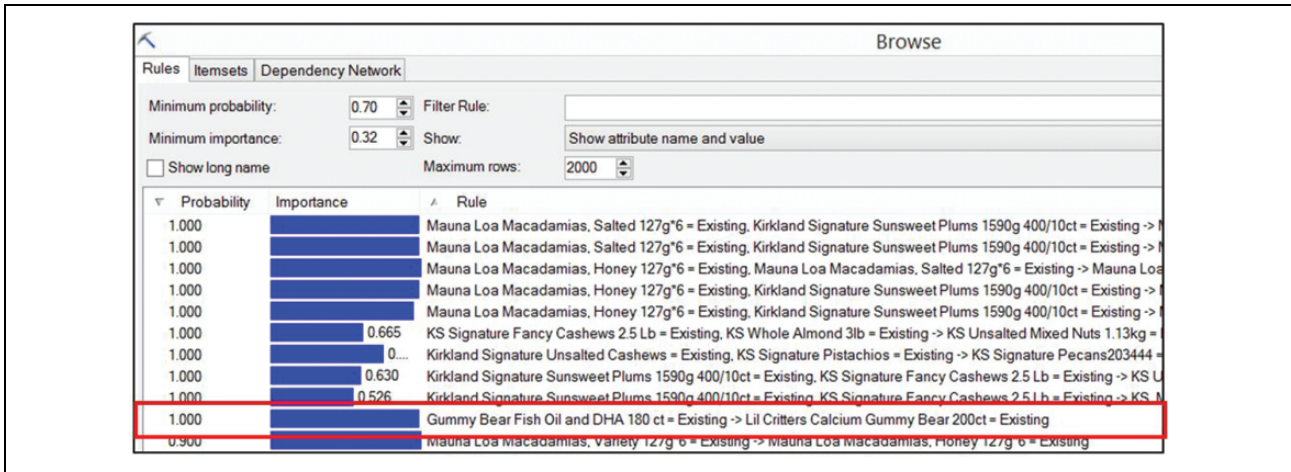


Figure 8. Rules generated by the “associate wizard.”

Table 6. KPIs records concerning the implementation of CLAS.

KPIs	Before CLAS implementation	After CLAS implementation	Percentage of improvement
SKU allocation time	51 min	45 min	11.8
Order-picking time	47 min	35 min	25.5
Product disposal rate	0.0067%	0.0049%	26.9
Product return/exchange rate	2.3%	2.0%	13.0

KPI: key performance indicator; CLAS: cloud-based location assignment system; SKU: stock-keeping unit.

that the system is able to identify SKUs with a potentially high turnover velocity, in turn allocating proper storage locations to them to shorten the traveling distance. The performance of these two KPIs reflects that CLAS is able to accelerate the storage-related operations of a packaged food warehouse.

ii. Reduction in loss caused by defective products

Besides enhancing efficiency of the warehouse, another important purpose of applying CLAS is to maintain the packaged food in a good state while the food is in the warehouse. This purpose is reviewed by the KPIs “product disposal rate” and “product return/exchange rate.” The records of the former KPI show that the ratio of defective products for a batch of newly arrived products decreased by 26.9%. This improvement implies that vulnerable products can be identified by CLAS and consequently be assigned to locations that better-suit their storage requirements.

The product return/exchange rate calculates the ratio of the returned or exchanged quantity to the sold quantity of all items in a month, where the reason for product return or

exchange is the discovery of product damage. The product return/exchange rate has a slight improvement of 13%. The improvement originates from the lower product disposal rate. A lower product disposal rate means the absolute amount of defective products is smaller, so the number of problematic products for the customer is thus smaller too. However, the results are not as significant as other KPIs because defective products are often hidden in the inner carton of the pallets, which cannot be detected, with or without the system. Nonetheless, these two KPIs prove that CLAS is able to reduce the loss related to defective products for the packaged food warehouse.

iii. Offering reference to range of SKU storage time

The methodology of using FARM in VS and range of storage time prediction in the VSM is a novel yet practical approach in the field. The fuzzy component of FARM enables a description of the parameters, which gives a more detailed explanation of the relationship between the parameters. Meanwhile, besides the confidence and support count thresholds, the discrepancy threshold further strengthens the rule selection process. Therefore, only rules that have strong prediction power are kept in the system for further usage. The predicted range of SKU storage time can be important reference for the warehouse dealing with e-orders, which involve uncertain outbound time.

The three types of benefits obtained through CLAS application indicate that the system can serve its purpose of assisting the decision maker of the packaged food warehouse, in turn improving the operations of the warehouse.

Cost analysis of CLAS implementation

After validating CLAS in terms of operational performance, its cost-effectiveness is evaluated in this section through conducting a cost analysis. The results of the analysis are summarized in Table 7.

Table 7. Cost analysis of CLAS implementation.

	Implementation cost	Before implementation	After implementation
System development and configuration (labor costs of two full-time system developers)	HK\$84,000	—	—
Maintenance cost per month (labor cost for a system operator and cost for employing cloud infrastructure)	—	—	HK\$17,000
Average labor cost for order-picking per month	—	HK\$50,000	HK\$35,000
Average cost incurred in product disposal and product missing per month	—	HK\$43,000	HK\$26,000

CLAS: cloud-based location assignment system.

In Table 7, the total cost of implementing the system is around HK\$84,000, which is the cost for system development and configuration. Meanwhile, the maintenance cost and the operational costs involved before and after implementing the system, that is, average labor costs for order picking, and average costs incurred in product disposal and loss per month are listed in the table. The saving brought by CLAS per year can thus be calculated as

$$\begin{aligned}
 & [(Average\ labour\ cost\ for\ order - picking\ per\ month \\
 & before\ implementation - Average\ labour\ cost\ for\ order - \\
 & picking\ per\ month\ after\ implementation) + (Average\ cost \\
 & incurred\ in\ product\ disposal\ and\ product\ missing\ per \\
 & month\ before\ implementation - Average\ cost\ incurred \\
 & in\ product\ disposal\ and\ product\ missing\ per\ month\ after \\
 & implementation) - Maintenance\ cost\ per\ month\ after \\
 & implementation] \times 12 \\
 & = HK\$ [(50,000 - 35,000) + (43,000 - 26,000) \\
 & - 17,000] \times 12 \\
 & = HK\$ 180,000.
 \end{aligned}$$

According to the calculated saving per year, the expected break-even point of CLAS implementation is

$$\begin{aligned}
 & HK\$ (\text{Money invested/saving per year}) \\
 & = HK\$ (84,000/180,000 \text{ per year}) \approx 0.5 \text{ years.}
 \end{aligned}$$

Based on the calculation, it takes around 0.5 years to justify the investment in CLAS. However, since the system will bring continuous savings to the warehouse as the system is continuously improved, it is likely that the proportion of the costs to the savings would be less than expected. Subsequently, the company can justify the investment in less than 0.5 years.

Conclusions

The trend of e-retailing is creating challenges to the e-fulfillment packaged food warehouse, which needs to operate at an even higher efficiency and maintain food in a good state. Since storage policy is closely related to the speed of receiving operations and order picking, as well as the storage conditions of food in the warehouse, a DSS that

is aimed to provide proper recommendations in SLAP is proposed. The DSS, CLAS, applies a novel FARM approach to identify variables of SKU storage time. The predicted range of SKU storage time is thus obtained and becomes a reference to the potential outbound time of products. In addition, the system integrates FL and ARM to generate allocation guidelines according to the turnover velocity, storage requirements, and ordering patterns of the packaged food. The system is proven to be feasible in the case study. The results of the evaluation of the system indicate that the system can improve the efficiency of receiving and order-picking operations, as well as reduce the loss incurred due to defective products in a packaged food warehouse. In addition, the costs of implementing CLAS can be recovered in 0.5 years. Therefore, by applying CLAS, a packaged food warehouse can be prepared itself to cope with the emerging unpredictable and time-sensitive e-retailing business environment.

In order to contribute to the related research areas in future, methodology and application of CLAS can be further studied. In terms of the methodology, input variables are currently weighted equally in the fuzzy inference engine. However, in the real situation, some variables could be more important than the others. In future, the relative importance of different factors should be studied and then be reflected in the engine through adjusting the weighting of the variables. Regarding the application of CLAS, since the trend of e-retailing has been observed in many other industries, such as home appliance and fashion industry, the storage operations concerning these industries would face similar challenges as the packaged food industry. To assist them, the parameters and variables of CLAS can be adjusted according to the storage requirements of their products. With more studies and applications, CLAS can be further enhanced to improve storage operations in other industries.

Authors' note

KH Leung is currently affiliated to Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors would like to thank the Research Office of the Hong Kong Polytechnic University for supporting this project (project code: RU5M).

Reference

- Verma V, Sharma D, and Sheth J. Does relationship marketing matter in online retailing? A meta-analytic approach. *J Acad Market Sci* 2016; 44: 206–217.
- Barnes C. Omni-channel retail — challenges and opportunities for packaging innovation. In: Burgess P (ed.) *Integrating the packaging and product experience in food and beverages: a road-map to consumer satisfaction*. Duxford: Woodhead Publishing, 2016, pp. 59–76.
- Sharma P, Nair V, and Jyotishi A. Patterns of online grocery shopping in India: an empirical study. In: *2014 International conference on interdisciplinary advances in applied computing*, Article No. 29, Amritapuri, India, 10–11 October 2014. New York: ACM. DOI: 10.1145/2660859.2660939.
- Nepomuceno MV, Laroche M, and Richard MO. How to reduce perceived risk when buying online: the interactions between intangibility, product knowledge, brand familiarity, privacy and security concerns. *J Retailing Consum Serv* 2014; 21(4): 619–629.
- Vanelander T, Deketele L, and Van Hove D. Commonly used e-commerce supply chains for fast moving consumer goods: comparison and suggestions for improvement. *A Leading Journal of Supply Chain Management. Intl J Logistics Res Apps* 2013; 16(3): 243–256.
- Subramanian N, Gunasekaran A, Yu J, et al. Customer satisfaction and competitiveness in the Chinese e-retailing: structural equation modeling (SEM) approach to identify the role of quality factors. *Expert Syst Appl* 2014; 41(1): 69–80.
- Pereira de Abreu DA, Cruz JM, and Paseiro LP. Active and intelligent packaging for the food industry. *Food Rev Int* 2012; 28(2): 146–187.
- Mkansi M, Qi B, and Green G. *A literature review on management of supply and distribution in the UK Online grocery retailing: a resource-based view*. In: *UK academy for information systems conference*, Article No. 32, Oxford, UK, 11–13 April 2011.
- Accorsi R, Manzini R, and Bortolini MA. Hierarchical procedure for storage allocation and assignment within an order-picking system. A case study. *Intl J Logistics Res Apps* 2012; 15(6): 351–364.
- Ene S and Öztürk N. Storage location assignment and order picking optimization in the automotive industry. *Int J Adv Manuf Technol* 2012; 60(5–8): 787–797.
- Chen MC and Wu HP. An association-based clustering approach to order batching considering customer demand patterns. *Omega* 2005; 33(4): 333–343.
- Chan FTS and Chan HK. Improving the productivity of order picking of a manual-pick and multi-level rack distribution warehouse through the implementation of class-based storage. *Expert Syst Appl* 2011; 38(3): 2686–2700.
- Pereira de Abreu DA, Cruz JM, and Paseiro LP(2012). Active and intelligent packaging for the food industry. *Food Rev Int* 2012; 28(2): 146–187.
- Amin ST and Alrub AMA. The role of MKIS in decision making: a conceptual framework. *Interdiscipl J Contemporary Res Bus* 2014; 5(9): 551–569.
- Obeidat M, North M, Richardson R, et al. Business intelligence technology, applications, and trends. *Int Manage Rev* 2015; 11(2): 47.
- Demirkan H and Delen D. Leveraging the capabilities of service-oriented decision support systems: putting analytics and big data in cloud. *Decis Support Syst* 2013; 55(1): 412–421.
- Hsieh TC, Yang KC, Yang C, et al. (2013) Urban and rural differences: multilevel latent class analysis of online activities and e-payment behavior patterns. *Internet Res* 2013; 23(2): 204–228.
- Guo ZX, Wong WK, and Guo C. A cloud-based intelligent decision-making system for order tracking and allocation in apparel manufacturing. *Int J Prod Res* 2014; 52(4): 1100–1115.
- Mann KS and Kaur N. Cloud-deployable health data mining using secured framework for clinical decision support system. In: *International conference and workshop on computing and communication (IEMCON)*, Vancouver, Canada, 15–17 October 2015, pp. 1–6. IEEE.
- Dong M, Li H, Ota K, et al. Multicloud-based evacuation services for emergency management. *IEEE Cloud Comput* 2014; 1(4): 50–59.
- Wu CH, Ho GTS, Lam CHY, et al. An online niche-market tour identification system for the travel and tourism industry. *Internet Res* 2016; 26(1): 167–185.
- Huang ML, Hung YH, and Chen WY. Neural network classifier with entropy based feature selection on breast cancer diagnosis. *J Med Syst* 2010; 34(5): 865–873.
- Liu H, Sun J, Liu L, et al. Feature selection with dynamic mutual information. *Pattern Recogn* 2009; 42(7): 1330–1339.
- Ząbkowski TS and Szczesny W. Insolvency modeling in the cellular telecommunication industry. *Expert Syst Appl* 2012; 39(8): 6879–6886.
- Elwakil E and Zayed T. Construction knowledge discovery system using fuzzy approach. *Can J Civil Eng* 2014; 42(1): 22–32.
- Kim BS and Smith JS. Slotting methodology using correlated improvement for a zone-based carton picking distribution system. *Comput Ind Eng* 2012; 62(1): 286–295.
- Wuthisirisart P, Noble JS, and Chang CA. A two-phased heuristic for relation-based item location. *Comput Ind Eng* 2015; 82: 94–102.
- Nettleton D. *Commercial data mining: processing, analysis and modeling for predictive analytics projects*. Amsterdam: Morgan Kaufmann, 2014.
- Olson DL, Delen D, and Meng Y. Comparative analysis of data mining methods for bankruptcy prediction. *Decis Support Syst* 2012; 52(2): 464–473.

30. Ma CY and Wang XZ. Inductive data mining based on genetic programming: automatic generation of decision trees from data for process historical data analysis. *Comput Chem Eng* 2009; 33(10): 1602–1616.
31. Sotoca JM and Pla F. Supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recogn* 2010; 43(6): 2068–2081.
32. Cho S, Hong H, and Ha BC. A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: for bankruptcy prediction. *Expert Syst Appl* 2010; 37(4): 3482–3488.
33. Pach FP, Szeifert F, Nemeth S, et al. Fuzzy association rule mining for data driven analysis of dynamical systems. *Hung J Ind Chem* 2005; 33 (1–2): 57–67.
34. Ho GTS, Ip WH, Wu CH, et al. Using a fuzzy association rule mining approach to identify the financial data association. *Expert Syst Appl* 2012; 39(10): 9054–9063.
35. Lee CKH, Tse YK, Ho GTS, et al. Fuzzy association rule mining for fashion product development. *Ind Manage Data Syst* 2015; 115(2): 383–399.
36. Hui YYY, Choy KL, Ho GTS, et al. *A Fuzzy association rule mining framework for variables selection concerning the storage time of packaged food*. In: *2016 IEEE international conference on fuzzy systems (FUZZ)*, Vancouver, Canada, 24–29 July 2016, pp. 671–677. IEEE.
37. Fouad RH, Samhoury MS, and Qamar AM. An intelligent preventive maintenance scheduling in ERP systems: a fuzzy logic approach. *Int J Adv Comput Technol* 2012; 4(23): 651–661.
38. Gill A. A fuzzy set theoretic approach to warehouse storage decisions in supply chains. In: Kahraman C and Öztayşi B (eds) *Supply chain management under fuzziness*. Berlin Heidelberg: Springer, 2014, pp. 471–488.
39. Ho GTS, Lau HCW, Choy KL, et al. Providing decision support for replenishment operations using a genetic algorithms based fuzzy system. *Expert Syst* 2015; 32(1): 23–38.
40. Chirici L and Wang K. A ‘lean’ fuzzy rule to speed-up a Taylor-made warehouse management process. In: Kovács GL and Kochan D (eds) *Digital product and process development systems*. Berlin Heidelberg: Springer, 2013, pp. 61–72.
41. Choy KL, Ho GTS, and Lee CKH. A RFID-based storage assignment system for enhancing the efficiency of order picking. *J Intell Manuf* 2014; 1–19. DOI: 10.1007/s10845-014-0965-9.
42. Li XL and Liu B. Rule-based classification. In: Aggarwal CC (ed.) *Data classification: algorithms and applications*. Boca Raton: CRC Press, 2014, pp. 121–156.
43. Chiang DMH, Lin CP, and Chen MC. The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres. *Enterprise Inf Syst* 2011; 5(2): 219–234.
44. Chiang DMH, Lin CP, and Chen MC (2014) Data mining based storage assignment heuristics for travel distance reduction. *Expert Syst* 31(1): 81–90.
45. Jain N and Srivastava V. Data mining techniques: a survey paper. *Int J Res Eng Technol* 2013; 2(11): 2319–1163.